Claims

[c1] 1. A method of automatically labeling an speech signal with phonic symbols for correcting pronunciation, comprising:
A step of establishing a phoneme-feature database, including using sample sound signal to establish a plurality of phoneme clusters;
A step of phonic symbol labeling, comprising:
Partitioning one sound signal into a plurality of frames, and calculating a feature set for each frame; and
Determining the phoneme cluster to which each frame belongs and labeling the frame with the corresponding phonic symbol; and
A step of pronunciation comparison, which compares the frames of two sound waves corresponding to the same phonic symbol or syllable, and perform grading and providing suggestion for improvement.

[c2] 2. The method according to Claim 1, wherein the step of establishing the phoneme- feature database further comprises analyzing the sample frames corresponding to each of the phoneme clusters.

[c3] 3. The method according to Claim 2, wherein the step of establishing the phoneme- feature database further comprises:
Recording sample sound signals;
Partitioning each sample sound signal into a plurality of sample frames;
Determining a phoneme cluster that each sample frame belongs to;
Calculating the feature set of each sample frame; and
Calculating the mean and variance of the feature sets of each phoneme cluster.

[c4] 4. The method according to Claim 2, further comprising the step of determining the phoneme cluster to which each frame belongs.

[c5] 5. The method according to Claim 2, wherein data contained in each phoneme cluster comprises the mean and variance of all the sample frames belong to the phoneme.

[c6] 6. The method according to Claim 1, wherein the step of phonic symbol labeling comprises:

Inputting a text string and a corresponding sound signal;
Looking up an electronic phonetic dictionary to find a string of phonic symbols that corresponds to the input text string;
Partitioning the input sound signal into a plurality of frames;
For each frame, calculating the probabilities that the frame belongs to different phonemes by comparing the frame's feature set against the data in the phoneme-feature database;
Obtaining an optimum labeling to frames that maximize the probability that the labeling is correct;
Displaying the phonic symbol corresponding to each frame.

[c7] 7. The method according to Claim 6, further comprising comparing the input text string and the corresponding input sound signal to obtain the label phonic symbol.

[c8] 8. The method according to Claim 6, when some of the phonic symbols corresponding to the input text string do not appear in the input sound signal, a normal operation is maintained, and other phonic symbols are used for labeling.

[c9] 9. The method according to Claim 6, when some intervals of the input sound signal contains silence, noise, or is redundant and does not correspond to any portion of the input text string, a normal operation is maintained, and other intervals of the sound signal are labeled.

[c10] 10. The method according to Claim 6, wherein the step of obtaining the optimum labeled phonic symbol includes a dynamic programming technique.

[c11] 11. The method according to Claim 10, wherein the dynamic programming technique includes using a comparison table, of which a row (or column) corresponds to a phonic symbol of the input phonic string, and a column (or row) corresponds to a frame in the input sound signal.

[c12] 12. The method according to Claim 11, wherein the step of obtaining the optimum labeling includes finding a path extending from upper left to lower right (or from lower right to upper left) which maximizes a predetermined utility

function (or minimizes a predetermined penalty function).

[c13] 13. The method according to Claim 1, wherein in the pronunciation comparison stage, one of the two sound signals is pre-recorded, and the other sound signal is recorded in real time.

[c14] 14. The method according to Claim 1, wherein the step of pronunciation comparison stage comprises comparing articulation accuracy, pitch, intensity and timing (rhythm).

[c15] 15. A user interface for automatically labeling speech signals with phonic symbols for correct pronunciation, comprising:
Waveform graphs, obtained by analyzing the sound signals;
Intensity variation graphs, obtained by analyzing the sound signals;
Pitch variation graphs, obtained by analyzing the sound signals;
Multiple pronunciation intervals on the waveform, intensity variation, and pitch variation graphs, where each interval corresponds to a phonic symbol and is bounded by two partitioning line segments; and
Phonic symbol labeling areas, which display the phonic symbols corresponding to the pronunciation intervals.

[c16] 16. The user interface according to Claim 15, where a user can select one or multiple adjacent pronunciation intervals and click a button or issue a command to replay the sound of those selected intervals.

[c17] 17. The user interface according to Claim 16, in which if one or more adjacent pronunciation intervals in the teacher's (or student's) speech signal are selected, the corresponding pronunciation intervals in the student's (or teacher's) speech signal will be selected automatically.

[c18] 18. A system for automatically labeling speech signals with phonic symbols to correct a language learner's pronunciation, comprising:
An input device, to input a text string and a corresponding sound signal;
An electronic phonetic dictionary, which is used to look up the string of phonic symbols that correspond to a text string;
An audio cutter that partitions the sound signals into multiple frames. The

frames may be overlapping;

A feature extractor, which extract a set of features from each frame;

A phoneme-feature database, including multiple phoneme clusters, where each of the phoneme clusters corresponds to a phonic symbol;

A phonic symbol labeler, which labels intervals of a speech signal with phonic symbols; and

An output device, which displays a waveform graph, a pitch variation graph, an intensity variation graph and phonic symbols corresponding to each pronunciation interval of the input sound signals.